Supporting Information

1- Variation of angles relevant to the coarse grain model as a function of the glycosidic torsion angles of α-maltose.

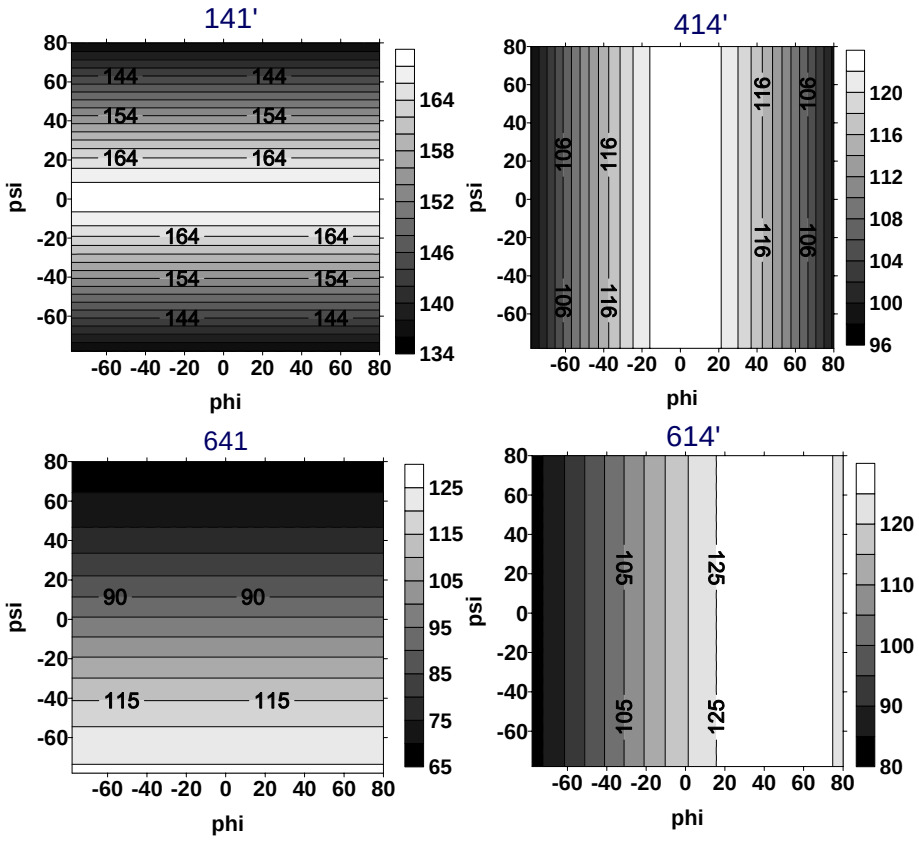

Figure 1Sa: Contour maps of the angles defined by carbons C1, C4, C6, C1', C6' and C4' of the atomistic DP2 shown in Figure 1 of the article as a function of the glycosidic torsion angles ϕ and ψ. The maps correspond to the rotation of the glycosidic torsion angles in an otherwise rigid α-maltose whose conformation was taken from the X-ray crystal structure. [1]

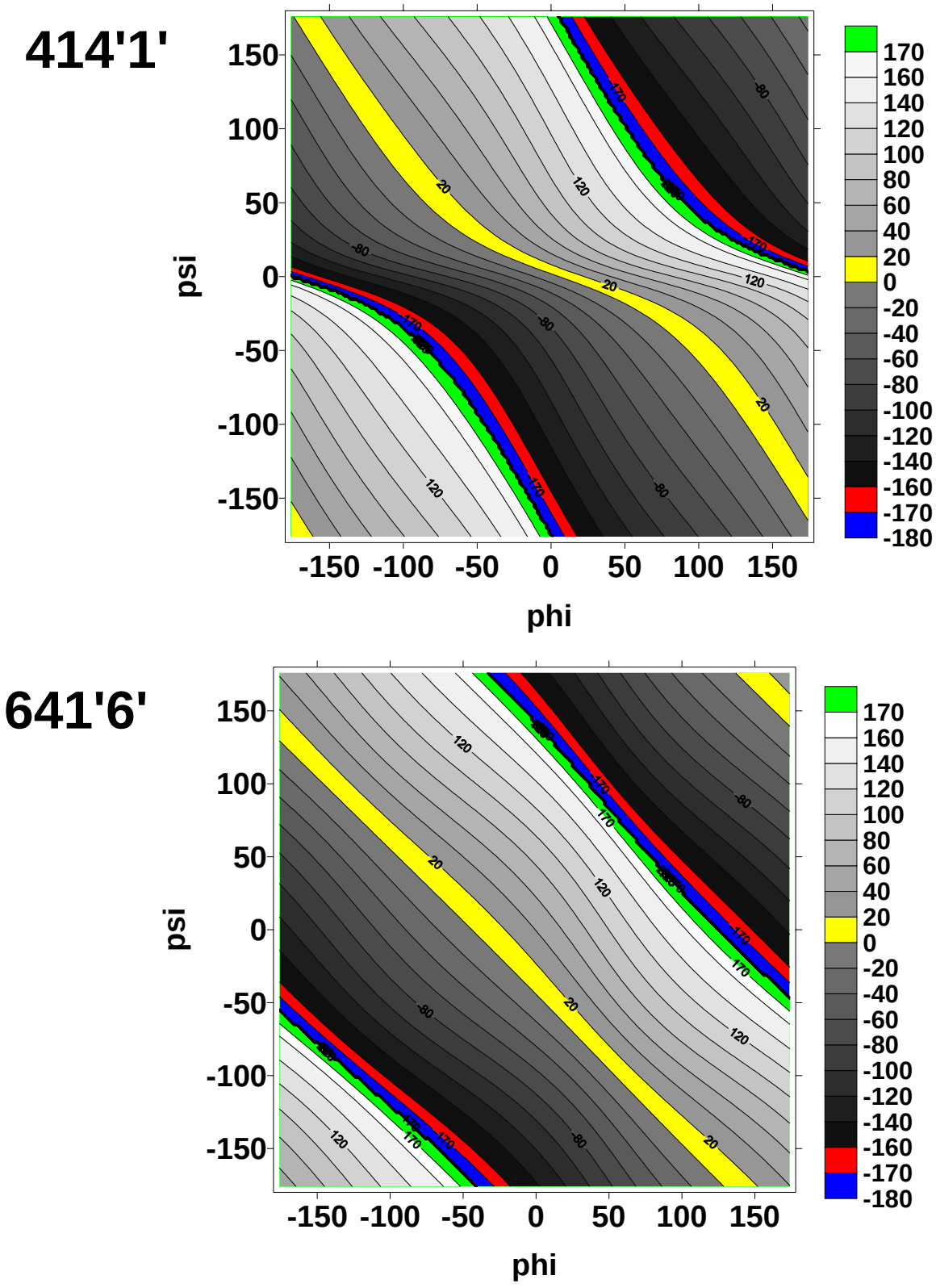

Figure 1Sb: Contour maps of the torsional angles defined by carbons C1, C4, C6, C1', C6' and C4' of the atomistic DP2 shown in Figure 1 of the article as a function of the

glycosidic torsion angles ϕ and ψ. The maps correspond to the rotation of the glycosidic torsion angles in an otherwise rigid α-maltose whose conformation was taken from the X-ray crystal structure. [1]

2-M3B parameterization:

In what follows, we present the details of the different steps of the optimization and discuss their results.

Step 1: Raw estimation of the nonbonding interaction potentials. The nonbonding interactions between two beads are represented by Morse potentials

$$V(R_{ij}) = D_o\{(e^{-0.5\alpha(R_{ij}/R_o-1)})^2 - 2(e^{-0.5\alpha(R_{ij}/R_o-1)})\} \qquad \text{Equation 1S}$$

The Morse potential accounts not only for the hydrogen bonded and dispersive (van der Waals) interactions of the atomistic description, but also the screened effect of the electrostatic interactions (the super-atoms bear no charges).

In the *M3B* model, the intermolecular energy for a pair of glucose molecules can be expressed as a sum of 9 Morse terms between the 3 beads of each *M3B* glucose. If, as a first approximation, we assume that all the beads have the same nonbonding parameters, we can estimate those parameters from the analysis of the intermolecular energy of a pair of α-glucose pairs at different distances. Due to the anisotropy of the glucose molecule, a sampling over a significant number of relative orientations for each intermolecular distance is needed. The glassy structures of α-glucose constitute an excellent source of pairs of molecules in a wide range of distances and relevant orientations, hence we computed the intermolecular energy and center of mass distance for the 2480 α-glucose pairs of the 5 amorphous glucose cells minimized at p=0 (see above) with the atomistic force field. The sorted energy results were averaged every 10 points to decrease the noise.

The resultant energy *vs* distance curve can be adjusted with a Morse potential function, as shown in Figure 2S. We obtained a first guess of the nonbonding parameters, assuming that the size of the beads coincides with the center of mass distance of minimum intermolecular energy in the atomistic model,

$$R_0^{M3B} = R_0^{A,INTERMOLECULAR} \qquad \text{Equation 2S}$$

and their potential well correspond to the D_0 of the intermolecular atomistic D_0 divided by the number of bead-bead interactions in two glucose system,

$$D_0^{M3B} = \frac{D_0^{A,INTERMOLECULAR}}{9} \qquad \text{Equation 3S}$$

The initial values of the Morse parameters estimated through this procedure were: R_o=5.1 Å, D_o=1.6 kcal/mol, and α=7.5.

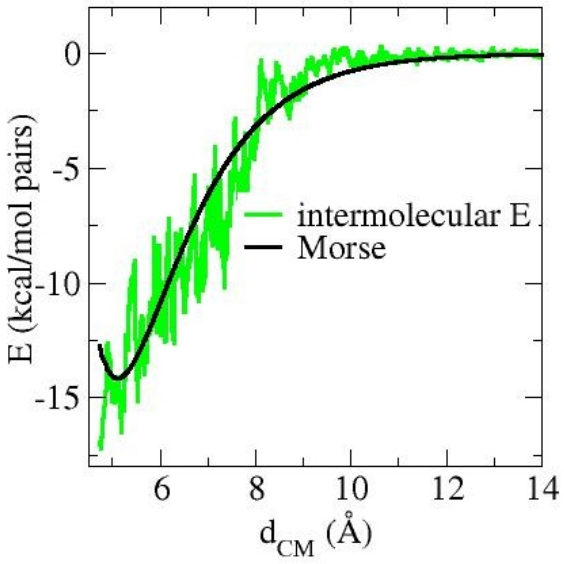

Figure 2S: Intermolecular energy a-glucose pairs (gray), as a function of the distance between the molecules center of mass. The original 2480 points were averaged (av. Length 10) to smooth the dispersion in energies for a same d that result from unfavorable orientations. The black curve

corresponds to a fit of the smoothed energy vs distance results with a Morse function. The morse parameters were D_0 = 14.2 kcal/mol, R_0=5.1 Å and α=7.5.

Step2: Simulated annealing refinement of the nonbonding parameters. In the previous step, we estimated a set of Morse parameters based on the intermolecular energy of a pair of α-glucose molecules as a function of distance. That first estimation assumes that all the three beads of *M3B* glucose have the same parameters. Starting with those values, we performed a simulated annealing search of a new set of Morse parameters. In this optimization, the structure was kept fixed and the nonbonding parameters changed until a cost function is minimized. The cost function measures how different are the *intermolecular* energies ε^A of the reference atomistic model and the intermolecular energy of the *same* configuration represented with the coarse grain model, ε^{M3B}

$$Cost(\{V_M\}) = \sum_{i=1}^{m}\sum_{p}\left(\frac{\varepsilon_{i,p}^{A} - \varepsilon_{i,p}^{M3B}(\{V_M\})}{\varepsilon_{i,p}^{A}}\right)^2 w_p \qquad \text{Equation 4S}$$

where the sum was over the *m* (5 for DP1 and DP4, and 3 for DP2) independent atomistic amorphous structures α-glucose minimized at pressure *p* in the range –1 to 20 GPa. The weights for the optimization w_p depended only on *p* and were chosen to give less significance to the extreme pressures: w_p was 1 for *p*= 0 to 3.5 GPa, 0.75 for –1 and 5 GPa structures and 0.5 otherwise.

The cost function depends exclusively on the Morse parameters of *M3B*, $\{V_M\}$. The parameters were changed in units of 0.5% of the values determined in Step1, and allowed to vary up to 50% for the R_{oi} and D_{oi}. The curvature parameter, α was allowed to vary between 7 and 13. It is expected that the beads will have a higher value of α (less soft) than the one obtained from fitting the interactions of the overall molecule.

The optimization was performed for the set of DP1, DP2 and DP4 bulk cells in the whole range of pressures. This optimization indicated that more than one set of parameters were required for the same super-atom, depending on the coordination of the monomer (m.c.) they belong to. The morse parameters for *B1*, for example, of a DP1 (m.c. 0) was not the same as the one for the reducing end of a DP2 (m.c. 1) nor for a non-terminal one in the same molecule (m.c. 2). The parameters depended on the monomer coordination, but not on the DP of the molecule: DP2 and DP4 parameters were transferable. The parameters obtained by the annealing procedure *are not* the final parameters of the optimization: they were obtained constraining the positions of the coarse grain particles. As the symmetry of the atomistic molecule and its *M3B* representation is not the same, the structure can relax to a lower energy configuration, changing the density, intermolecular energy and even the size of the molecules. The same Monte Carlo search scheme can be implemented performing minimizations of all *M3B* structures for each set of Morse parameters. The computational cost would be exceedingly high, however, and the valence parameters should also be included in the set of optimizing variables. Instead, we used the best Morse parameters obtained with the Monte Carlo Simulated Annealing optimization (not listed) along with the valence parameters from the step *V1* as an input for the final optimization of the force field.

Step3: Estimation of M3B valence potentials from single molecule atomistic simulations. The valence terms of the coarse grain potential should guarantee that the molecule sample the same configurational space in the atomistic and coarse grain representations. As the number of beads is lower than the number of atoms, our goal is to sample with the coarse grain model the configurational space defined by the atoms *C1*,

C4 and *C6* in the atomistic model. The use of atomistic distributions to obtain effective potentials for coarse grain particles has been already employed for the parameterization of polymers. [2,3] For the present model, the configurational space to sample is conformed by the microscopic distributions of distances, angles and torsions between *C1*, *C4* and *C6* in the atomistic model.

The idea of the procedure is to minimize the difference between the distribution *P* of a distance (bond, angle or torsion) in the atomistic model and the coarse grain one,

$$Cost(\{V\}) = \int_l dl P^{A,T,n}(l) - P^{M3B,T,n}(l,\{V\}) \qquad \text{Equation 5S}$$

where *l* is the value of the measurement, *{V}* denotes the set of *M3B* valence parameters (*e.g. k* and r_o of the bond between *C1* and *C4*), $P^{A,T,n}(l)$ is the distribution function of the term *n* in the atomistic model at a temperature *T*, and $P^{M3B,T,n}(l,\{V\})$ is the distribution function for the same term, at the same *T* in the *M3B* model, evaluated with the parameter set *{V}*. If the distributions are not strongly coupled, the function $P^{M3B,T,n}(l,\{V\})$ can be interpreted as the thermal distribution corresponding to the coarse grain potential energy for the corresponding valence term *n*,

$$P^{M3B,T,n}(l,\{V\}) \propto \exp\left(-\frac{E^{M3B,n}(l,\{V\})}{RT}\right) \qquad \text{Equation 6S}$$

where *R* is the gas constant and $E^{M3B,n}(\{V\})$ is the *M3B* energy expression for the bond, angle or torsion being analyzed.

The distances, angles and torsions of the *M3B* model have a direct correspondence with distances, angles and torsions between *nonbonded* atoms. Our procedure computes the distance distributions over atomistic NVT Molecular Dynamics runs from which we

obtain the initial set of valence parameters via the minimization of the partial cost function (equation 6S) of each term with respect to the corresponding valence term parameters. The minimization of the cost function was done with the software Grace.[4]

Bond parameters: For each of the reference atomistic simulations we computed the distributions of the intramonomer bead distances ($r_{ik} = |\vec{r}_{Ci} - \vec{r}_{Ck}|$ for *ij* pairs *14, 16, 46* of the same glucose monomer in the molecule) and the intermonomer bead distances ($r_{14'} = |\vec{r}_{C1} - \vec{r}_{C4'}|$ where the two *C* atoms correspond to neighbor glucose residues linked by a glycosidic bond).

Figure 3S shows those bonds distributions at 300 K. Each of the coarse grain bond type (*14, 16, 46* and *14'*), showed a characteristic distribution. There was no difference, however, for the same bond type in DP1, DP2 and DP4. All the coarse-grain bonds atomistic distributions were unimodal, thus confirming that a simple harmonic potential for the bonds was appropiate. Table 1S summarizes the values of the bond parameters that minimize the cost function

$$Cost^{n}(\{k, r_o\}) = \int_0^{\infty} dr P^{A,T,n}(r) - \frac{\sqrt{\pi} RT}{k} \exp\left(- \frac{k(r - r_o)^2}{2RT} \right) \qquad \text{Equation 7S}$$

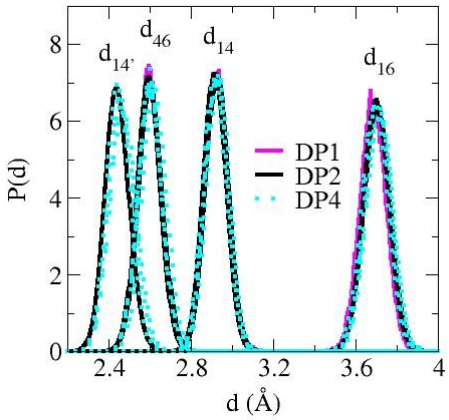

Figure 3S: *Bond distribution of the particles represented in the coarse grain model for the monomer, dimer and tetramer at 300 K. Length of the atomistic simulations: 2ns for DP1, 60 ns for DP2 and 20 ns for DP4.*

Table 1S: Bond parameters for the coarse grain model, derived from Boltzmann inversion of the gas phase bond distributions of the atomistic DP1, DP2 and DP4 at 300 K.

Bond type	R_0 *(Å)*	k_b *(kcal mol^{-1}Å^{-2})*
14	2.915	195
16	3.689	150
46	2.576	177
14'	2.48	176

The force constants for the coarse grain bonds obtained from the inversion of the atomistic distributions of gas phase simulations are approximately four times lower than the bond constants employed for the atomistic simulations, 700 kcal mol^{-1}Å^{-2}. The lower constant and higher mass of the beads with respect to the atoms rendered a significant decrease in the frequency of the bonds. If we assume that mass of *B6* corresponds

approximately to that of the exocyclic group (30 a.m.u.) and the rest of the mass is distributed equally between the ring components B1 and B4 (75 a.m.u. if terminal, 66 a.m.u. if not terminal, to account for the loss of water in the polymerization), the highest vibrational frequency of the bead bonds would be 1.12 10^{-13} s^{-1}, and thus a time step as large as 20 fs could be used for the simulations of this system, assuming 1 oscillation every 6 steps for an accurate integration of the movement equations.

Angle parameters: There are four different coarse grain angles θ to be parameterized in the *M3B* force field (*141', 414', 641'* and *614'*). They involve beads of different monomers, because the valence interactions between coarse grain particles of the same glucose residue are completely determined by the three intra-residue bonds. The atomistic distributions of the angles between *C1*, *C4* and *C6* of neighbor monomers at 300 K are shown in Figures 4S.

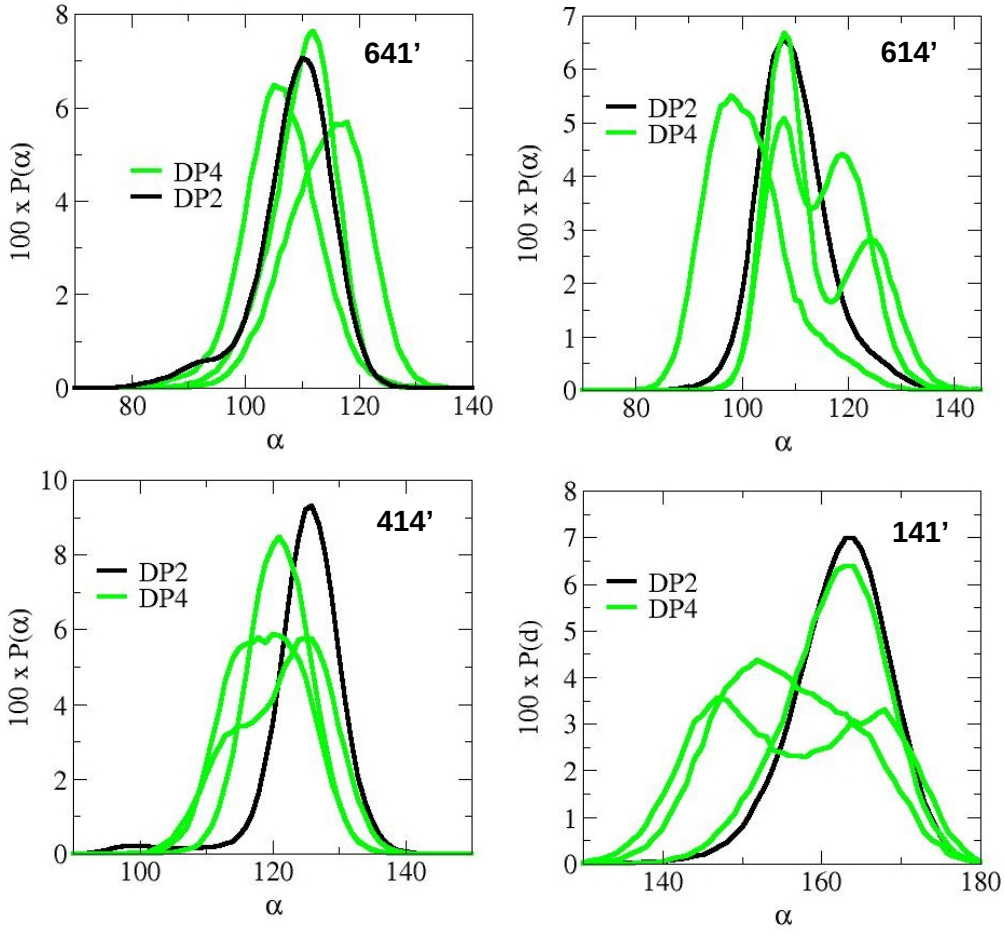

Figure 4S: Angle distribution of the angles between the carbon 641', 614', 414' and 141' for the atomistic DP2 and DP4 at 300 K. Each angle type occurs three times in the tetramer.

The distribution for the angles can be represented by a unimodal function, though it is more a smooth function for DP2 than DP4, for which the distribution is affected by nonbond interactions. For this reason, we obtained the parameters from DP2 distributions, assuming a harmonic potential for the bending angle (equation 5, main text), with which we minimized the cost function expressed by

$$Cost^{n}(\{k_\theta, \theta_o\}) = \int_0^{\pi} d\theta P^{A,T,n}(\theta) - \frac{2RT}{k_\theta} erf(\pi) \exp\left(-\frac{k_\theta(\theta-\theta_o)^2}{2RT}\right) \qquad \text{Equation 8S}$$

The resultant optimized parameters are shown in Table 2Sa.

Table 2Sa: *bending angle parameters of the coarse grain force field, obtained atomistic* gas phase *distributions for DP2 at 300 K.*

Angle type	θ_0^{V1} *(degrees)*	K_θ^{V1} *(kcal mol^{-1} radian^{-2})*
141'	163	60
414'	126	115
641'	110	66
614'	110	56

Table 2Sb shows the coarse grain force field angle parameters estimated from the distributions of the DP4 molecule shown in Figure 4S.

Table 2Sb: *bending angle parameters of the coarse grain force field, obtained atomistic* gas phase *distributions for DP4 at 300 K.*

Angle type	θ_0^{V1} *(degrees)*	K_θ^{V1} *(kcal mol^{-1} radian^{-2})*
141'	142	25
414'	117	65
641'	113	20
614'	113	20

Torsion parameters: From the atomistic simulations of DP2 and DP4 we computed the distribution of all coarse grain torsional angles φ involving the carbons *C1, C4* and *C6* of each monomer. The inter-monomer *14'* coarse grain bond is the only rotatable one between two monomers in a linear chain. Four different torsions can be written around that bond (*614'6*, *141'4, 614'1* and *414'6*). For fixed values of the coarse grain angles, these torsions are not independent. However, as the angles are allowed to vary during the dyanamics, we parameterized them independently. Figure 5S shows the distribution of torsional angles for *641'6'* and *414'1'* for the atomistic DP2. Besides the torsion around *14'* bond, we can write torsional terms for the non-rotatable bonds (*4614'*, *1641'*, *6141'* and *6414'*). In addition, there is a coarse grain torsion involving three monomers, *41'4'1''*. All the possible torsions were parameterized for *M3B*. We have analyzed the distribution for all possible coarse grain torsion angle to obtain the effective torsional

potential. For simplicity, we will illustrate with the results for the *141'4'* and *641'6'* rotatable bond torsions.

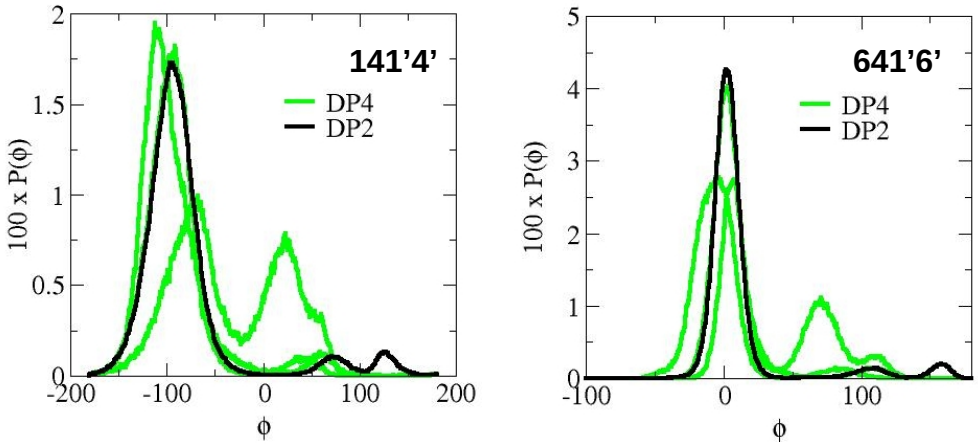

Figure 5S: Torsion angle distribution for the atomistic dimer and tetramer. 414'1' (panel A) and 641'6' (panel B). Results obtained from NVT simulations of the isolated molecules at T = 300 K during 60 ns for DP2 and 20 ns for DP4.

There are two main states for the *641'6* and *141'4'* torsions in the atomistic simulation of DP2, as can be more clearly seen from the analysis of the effective potential,

$$\ln\left(P^{A,T,n}(\phi_{ijkl})\right) \propto \exp\left(-\frac{E(\phi_{ijkl})}{RT}\right) \qquad \text{Equation 9S}$$

shown in Figure 6S). The higher energy conformational state is split into two subsets. Figure 7S shows the population of conformational states (i.e. Ramachandran plot) for atomistic DP2 at 300 K as a function of both the atomistic glycosidic torsion angles [ϕ, ψ] and the *M3B*-torsions [$\varphi_{414'1'}$, $\varphi_{641'6'}$]. The two main conformational states are encircled with solid and dashed lines. The population of conformational states visited by atomistic maltose in this 60 ns NVT simulation shows good agreement with the conformational

map for this molecule determined by Schmidt *et. al* using umbrella sampling and CHARMM force field.[5].

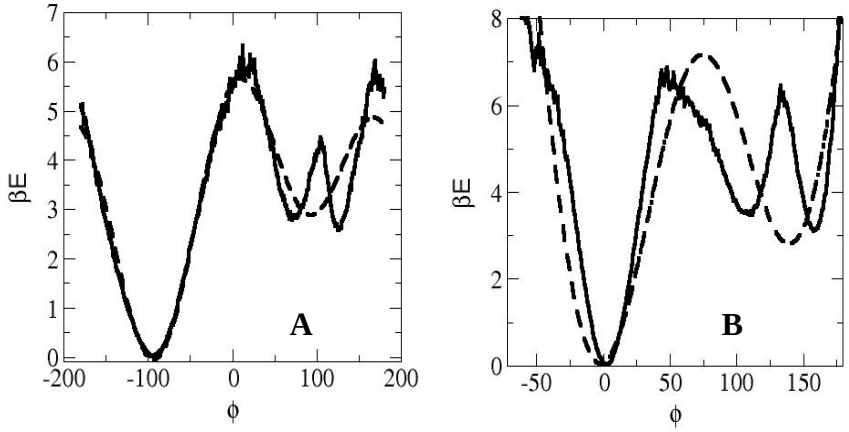

Figure 6S: Effective torsional potential, E. A = 4141 B = 6416. Dashed lines correspond to the shift dihedral functions with the values of Table 3S. $\beta=(RT)^{-1}$. . Atomistic simulation of α-maltose, T= 300 K.

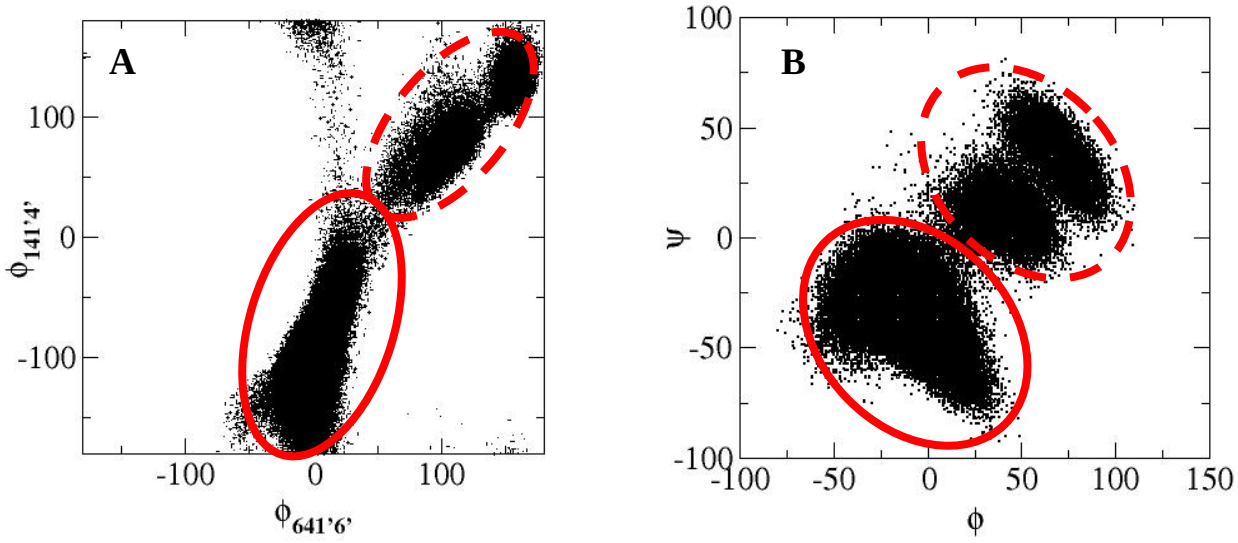

Figure 7S: Population of [$\phi_{141'4'}$,$\phi_{641'6'}$] torsional angles (A) and [ϕ ,ψ] atomistic torsional angles around the glycosidic bond (B) for a 60 ns NVT atomistic simulation of a α-maltose molecule at 300 K. The dashed and solid red lines encircle the same set of configurations in each of the bead-ramachandran (A) and standard-ramachandran (B) representations.

In the parameterization of the coarse grain torsional potential we first assumed that all the effective four body interactions are given by torsions. As done for the bond and angle terms, this leads to effective coarse grain torsion potentials as shown in Figure 6S (dashed lines). There is a contribution, however, of the nonbonding interactions. Thus – with all the current nonbond and valence parameters- we run NVT *coarse grain simulations* of a *M3B* maltose molecule at 300 K (integration steps 10 fs) and determined the resultant *coarse grain distribution*. The time step in the coarse grain simulations was 10 fs. The coarse grain distributions were compared with the atomistic ones, and the coarse grain torsional parameters modified while necessary to improve the matching. The process was done until the peaks of the coarse grain and atomistic distributions differ less than 20 degrees and (when the distributions were not unimodal) the relative height of the peaks was within 15% of the atomistic-simulation value. The resultant parameters for the coarse grain shift-dihedral potential (equation 6, main text) can be seen in Table 3S.

Table 3S: Torsion parameters for M3B (see equation 6, main text)

Torsion type	B_1 *(kcal mol^{-1})*	φ_1^o *(degrees)*	B_2 *(kcalmol^{-1})*	φ_2^o *(degrees)*	B_3 *(kcalmol^{1})*	φ_3^o *(degrees)*
414'1'	1.0	55	1.8	-20	--	--
641'6'	4.3	-135	2.8	130	--	--
4614'	30.0	-58	--	--	--	--
1641'	42.0	40	--	--	4.0	30
641'4'	3.6	93	2.5	-145	--	--
614'1'	1.0	27	2.1	-115	--	--
6141'	10.1	-65	1.3	71	--	--
6414'	15.7	112	4.3	-29	--	--
41'4'1''	15.0	-160	--	--	--	--

Step4: Final optimization of the coarse grain force field.

Up to this point we have determined coarse grain valence parameters that reproduce the vacuum distributions and nonbond parameters that optimize the energetics at fixed positions. Our aim is to obtain a set of parameters appropriate for constant pressure simulation bulk malto-oligosaccharides. The angle and bond constants of Tables 1S and 2S are considerably lower than those typical of atomic bonds and angles, indicating that these are soft modes. This would mean that the molecules themselves could be excessively compressed or expanded when different stresses are applied to the bulk systems. In effect, we have observed that the use of the bond and angle parameters shown

in Tables 2S and 3S lead to a non-physical response of the system to the applied pressure: the molecules are too compressible compared with the atomistic results. While the vacuum simulation sample only the softer modes of the molecule, the bulk compression "senses" the hard-core resistance of the glucose ring to be compacted.

On the other hand, the Morse parameters that resulting from Step 2 of the parameterization procedure correspond to a non-relaxed *M3B* geometry: the minimized atomistic cells were mapped into *M3B* and the energy computed without further relaxation of the beads or cell coordinates. As the symmetry of the atomistic and coarse grain model are not the same, a minimization of the bead coordinates may lead to rearrangements that change the position of the molecules and the density of the system. Not only the density could change, but also the intermolecular energy can be further lowered by the relaxation of the *M3B* model. Our strategy to correct these deficiencies of the *V1* and *NB2* set of parameters consisted of allowing the *M3B* model relaxation and re-optimize the bond, angles and Morse parameters (we assumed the torsions were not changed by compression) via the minimization of a cost function that considers the difference in density, intermolecular energy, cell parameters, angle and bond length and root mean square displacement between the atomistic model minimized with the atomistic force field and the mapped *M3B* model minimized with the coarse grain force field. To restrict the phase space of the Morse parameters, we started from the *NB2* parameters and rescaled the R_o for *all* the nonbond interactions by the ratio between the relaxed density and the density of the corresponding atomistic cell minimized in the same conditions:

$$R_{o,i}^{m+1} = R_{o,i}^{m} \frac{\rho_{M3B}^{m}}{\rho_{atom}} \qquad \text{Equation 10S}$$

A similar scaling was done for the Do parameters, from the comparison of the intermolecular energies in the same minimized atomistic and coarse grain cells

$$D_{o,i}^{m+1} = D_{o,i}^{m} \frac{E_{atom}}{E_{M3B}^{m}} \qquad \text{Equation 11S}$$

The values of α were scaled in an equivalent way comparing the compressibility of the atomistic and *M3B* models, in the studied range of pressures. The bonds and angle were refined in a similar manner: The new equilibrium positions were obtained from the mean value of the distribution at zero pressure in the condense phase and the bonds and angles constants were iteratively adjusted using the mean values of each bond type in the atomistic and M3B simulation:

$$k_i^{m+1} = k_i^{m} \frac{\left\langle d_{i\,atom} \right\rangle}{\left\langle d_{i\,M3B}^{m} \right\rangle} \qquad \text{Equation 12S}$$

We refer to the main text of the article to the results of these refinements.

References of the Supporting Information Section:

(1) Takusagawa, F.; Jacobson, R. A. *Acta Crystallographica Section B-Structural Science* **1978**, *34*, 213.

(2) Baschnagel, J.; Binder, K.; Paul, W.; Laso, M.; Suter, U. W.; Batoulis, I.; Jilge, W.; Burger, T. *Journal of Chemical Physics* **1991**, *95*, 6014.

(3) Tschop, W.; Kremer, K.; Batoulis, J.; Burger, T.; Hahn, O. *Acta Polymerica* **1998**, *49*, 61.

(4) Turner, P. J.; Team, G. D. Grace Portland, OR, 1996-2002.

(5) Schmidt, R. K.; Teo, B.; Brady, J. W. *Journal of Physical Chemistry* **1995**, *99*, 11339.